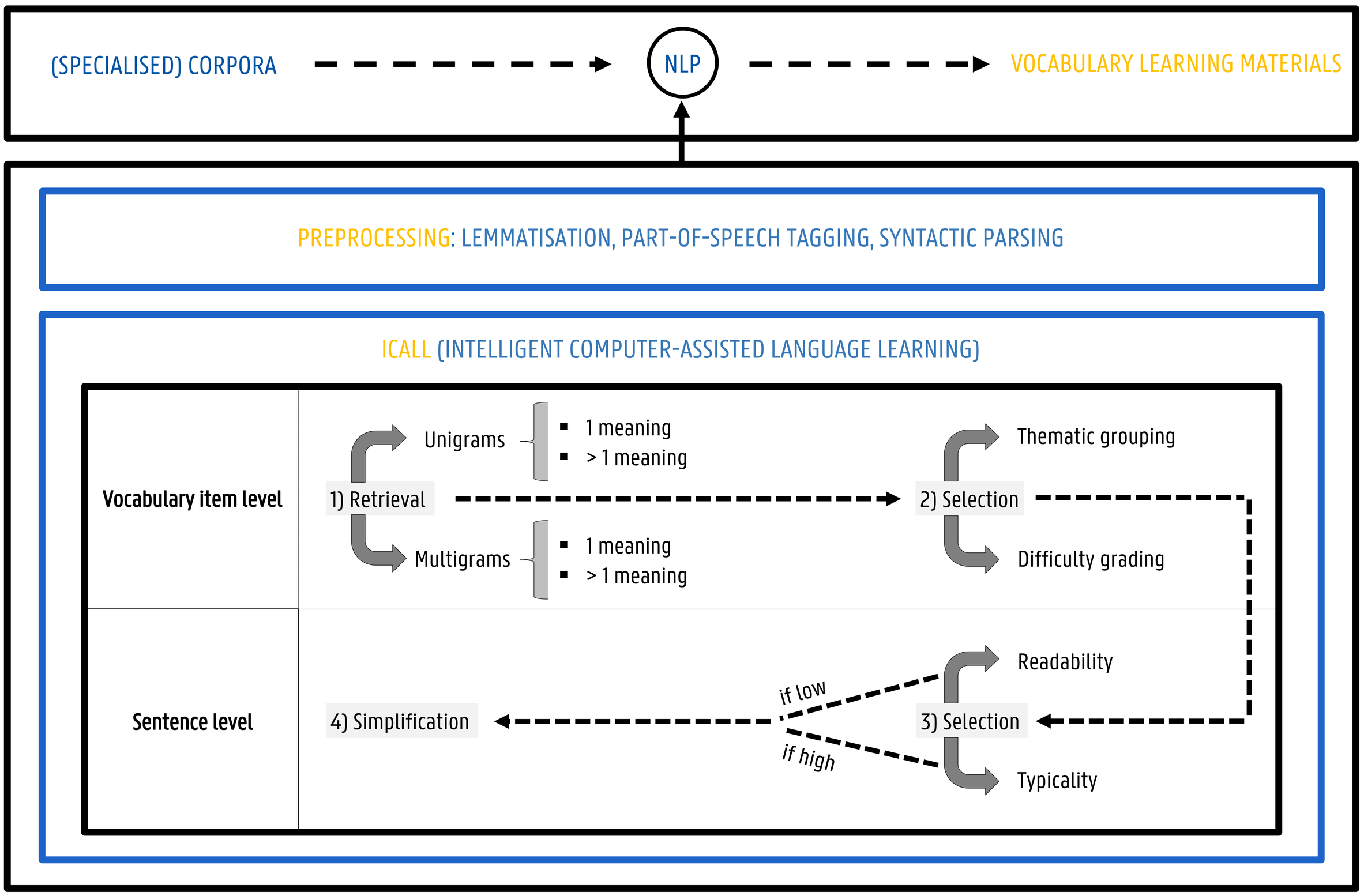
language and translation technology team

IVESS: INTELLIGENT VOCABULARY AND EXAMPLE SELECTION FOR SPANISH VOCABULARY LEARNING

JASPER DEGRAEUWE & PATRICK GOETHALS

NLP-ASSISTED GENERATION OF VOCABULARY LEARNING MATERIALS



RESEARCH OBJECTIVES

- Improving specific aspects of preprocessing (e.g. lemmatisation and syntactic parsing of pronominal verbs)
- Improving existing ICALL techniques and methodologies for creating vocabulary learning materials for Spanish as a foreign language (SFL)
  - Supervised word sense disambiguation with collection of sense-annotated data integrated in online vocabulary exercises
  - Automatic thematic selection of vocabulary items for a specialised domain with "keyness" metrics[1]
  - Adaptation of example sentence selection methodology for Swedish[2] to Spanish
- Gauging attitudes of students and teachers towards working with NLP-generated vocabulary learning materials
  - Questionnaires
  - Think-aloud interviews
- Development of didactically motivated crowdsourcing

PRELIMINARY RESULTS

- Reannotation proposal for potentially reflexive pronouns in Spanish Universal Dependencies treebanks[3]
- Machine learning classifier for automatic difficulty grading (non-disambiguated unigrams)[4]
  - 62% accuracy
  - Features used to predict difficulty level: frequency (in percentiles), cognateness and appearance in existing SFL vocabulary learning methods
- Comparison of automatic thematic vocabulary selection (non-disambiguated unigrams) with teacher judgements[5]
  - 72% overlap on average
  - Substantial agreement between automatic selection and teacher judgements (Cohen's kappa = 0,61)

GHENT UNIVERSITY

REFERENCES
[1] Gabrielatos, C. (2018). Keyness analysis: nature, metrics and techniques. In Taylor, C. & Marchi, A. (Eds.) *Corpus approaches to discourse: a critical review* (pp. 225-258). Oxford: Routledge.
[2] Pilán, I. (2018). *Automatic proficiency level prediction for Intelligent Computer-Assisted Language Learning* (PhD thesis), University of Gothenburg.
[3] Degraeuwe, J. & Goethals, P. (subm.) Reflexive pronouns in Spanish Universal Dependencies. *Procesamiento del Lenguaje Natural*.
[4] Goethals, P., Tezcan, A. & Degraeuwe, J. (2019). Vocabulary selection for didactic purposes: report on a machine learning approach. *Argentinian Journal of Applied Linguistics*, *7*(2), 34-51.
[5] Degraeuwe, J. & Goethals, P. (subm.). La selección temática del vocabulario para fines didácticos: evaluación de un acercamiento cuantitativo. *Revista de Lingüística y Lenguas Aplicadas*.